Exhibit B-10

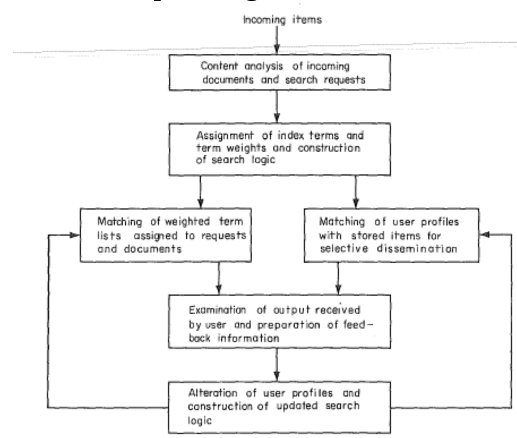
ACC - 10

Invalidity Chart
Salton '89 in view of Herz and Additional Prior Art References

Invalidity Chart
Salton ’89 in view of Herz and Additional Prior Art References

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
1. A data processing method for enabling a user utilizing a local computer system having a local data storage system to locate desired data from a plurality of data items stored in a remote data storage system in a remote computer system, the remote computer system being linked to the local computer system by a telecommunication link, the method comprising the steps of:	Salton ’89 p. 229 “Information retrieval systems process files of records and requests for information, and identify and retrieve from the files certain records in response to the information requests. The retrieval of particular records depends on the similarity between the records and the queries, which in turn is measured by comparing the values of certain attributes attached to records and information requests.”	Herz 79:11-14 “A method for cataloging a plurality of target objects that are stored on an electronic storage media, where users are connected via user terminals and bidirectional data communication connections to a target server that accesses said electronic storage media.” Herz 1:19-21 “This invention relates to customized electronic identification of desirable objects, such as news articles, in an electronic media environment.” Herz *See also* Abstract; 1:18-43; 4:35-48; 28:41–55:42; Figures 1-16.	Salton ‘68 p. 7 “Because of their special importance in the present context, it is useful to describe in more detail the operations that lead to the retrieval of stored information in answer to user search requests. In practice, searches often may be conducted by using author names or citations or titles as principal criteria. Such searches do not require a detailed content analysis of each item and are relatively easy to perform, provided that there is a unified system for generating and storing the bibliographic citations pertinent to each item.” Braden 5:2-6 “In accordance with our broad teachings, the present invention satisfies this need by employing natural language processing to improve the accuracy of a keyword-based document search performed by, e.g., a statistical web search engine.” Culliss 1:28-31 “Given the large amount of information available over the Internet, it is desirable to reduce this information down to a manageable number of articles which fit the needs of a particular user.” Ahn 1:31-33 “The present invention is directed to a system and method for searching through documents maintained in electronic form. The present invention is capable of searching through individual documents, or groups of documents.”

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			Brookes 1:9-14 "This invention relates to information technology and, in particular, to a method and apparatus whereby users of a database system may be alerted to important information including text, graphics and other electronically stored information within the system and by which means information may be efficiently disseminated." Dasan 1:10-15 "The present invention relates to information retrieval. More specifically, the present invention relates to a client server model for information retrieval based upon a user-defined profile, for example, for the generation of an "electronic" newspaper which contains information of interest to a particular user." Dedrick *See, e.g.*, Abstract, Figures 1-8. Krishnan *See* 1:6-12. Kupiec 3:23-29 "The present invention provides a method for answer extraction. A system operating according to this method accepts a natural-language input string such as a user supplied question and a set of relevant documents that are assumed to contain the answer to the question. In response, it generates answer hypotheses and finds these hypotheses within the documents." Reese 1:55-57 "A method and a system for requesting and retrieving information from distinct web network content sites is disclosed."

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			Menczer p. 157 "In this paper we discuss the use of algorithms based on adaptive, intelligent, autonomous, distributed populations of agents making local decisions as a way to automate the on-line information search and discovery process in the Web or similar environments." Armstrong p. 4 "We have experimented with a variety of representations that re-represent the arbitrary-length text associated with pages, links, and goals as a fixed-length feature vector. This idea is common within information retrieval systems [Salton and McGill, 1983]. It offers the advantage that the information in an arbitrary amount of text is summarized in a fixed length feature vector compatible with current machine learning methods."
(a) extracting, by one of the local computer system and the remote computer system, a user profile from user linguistic data previously provided by the user, said user data profile being representative of a first linguistic pattern of the said user linguistic data;	Salton '89 p. 405-6 "To help furnish semantic interpretations outside specialized or restricted environments, the existence of a *knowledge base* is often postulated. Such a knowledge base classifies the principal entities or concepts of interest and specifies certain relationships between the entities. [43-45] . . . . The literature includes a wide variety of different knowledge representations . . . [one of the] best-known knowledge-representation techniques [is] the *semantic-net*. . . . In generating a semantic network, it is necessary to decide on a method of representation for each entity, and to	Herz 56:19-27 "Initialize Users' Search Profile Sets. The news clipping service instantiates target profile interest summaries as search profile sets, so that a set of high interest search profiles is stored for each user. The search profiles associated with a given user change over time. As in any application involving search profiles, they can be initially determined for a new user (or explicitly altered by an existing user) by any of a	Salton '68 p. 9, Fig. 1-3  Fig. 1-3 Simplified user feedback process. "different content analysis procedures are available to generate identifiers for documents

The '067 Patent	Salton '89	Herz	Additional Prior Art References
	relate or characterize the entities. The following types of knowledge representations are recognized: [46-48]. . . . A linguistic level in which the elements are language specific and the links represent arbitrary relationships between concepts that exist in the area under consideration." Salton '89 p. 378 "A prescription for a complete language-analysis package might be based on the following components: A *knowledge base* consisting of stored entities and predicates, the latter used to characterize and relate the entities."	number of procedures, including the following preferred methods: (1) asking the user to specify search profiles directly by giving keywords and/or numeric attributes, (2) using copies of the profiles of target objects or target clusters that the user indicates are representative of his or her interest, (3) using a standard set of search profiles copied or otherwise determined from the search profile sets of people who are demographically similar to the user." Herz 6:58-60 "Each user's target profile interest summary is automatically updated on a continuing basis to reflect the user's changing interests." Herz 7:26-29 "The accuracy of this filtering system improves over time by noting which articles the user reads and by generating a measurement of the depth to which the user reads each article. This information is then used to update the user's target profile interest	and requests. . . statistical and syntactic procedures to identify relations between words and concepts, and phrase generating methods." Salton '68 p. 11 (Statistical association methods, Syntactic analysis methods, and Statistical phrase recognition methods) Salton '68 p. 33 "The phrase dictionaries. Both the regular and the stem thesauruses are based on entries corresponding either to single words or to single word stems. In attempting to perform a subject analysis of written text, it is possible, however, to go further by trying to locate phrases consisting of sets of words that are judged to be important in a given subject area." Salton '68 p. 35-36 "The syntactic phrase dictionary has a more complicated structure, as shown by the excerpt reproduced in Fig. 2-6. Here, each syntactic phrase, also known as criterion tree or criterion phrase, consists not only of a specification of the component concepts but also of syntactic indicators, as well as of syntactic relations that may obtain between the included concepts. . . . More specifically, there are four main classes of syntactic specifications, corresponding to noun phrases, subject-verb relations, verb-object relations, and subject-object relations." Braden 7:19-23 "Generally speaking and in accordance with our present invention, we have recognized that precision of a retrieval engine can be significantly enhanced by employing natural language processing to

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
		summary.” Herz 27:47-49 “[T]he disclosed method for determining topical interest through similarity requires users as well as target objects to have profiles.” Herz 27:62-67 “In a variation, each user’s user profile is subdivided into a set of long-term attributes, such as demographic characteristics, and a set of short-term attributes . . . such as the user’s textual and multiple-choice answers to questions” Herz 56:20-28 “As in any application involving search profiles, they can be initially determined for a new user (or explicitly altered by an existing user) by any of a number of procedures, including the following preferred methods: . . . (2) using copies of the profiles of target objects or target clusters that the user indicates are representative of his or her interest.” Herz 59:24-27 “The user’s desired attributes . . . would	process, i.e., specifically filter and rank, the records, i.e., ultimately the documents, provided by a search engine used therein.” Braden *See, e.g.*, 11:62-14:61. Culliss 3:46-48 “Inferring Personal Data Users can explicitly specify their own personal data, or it can be inferred from a history of their search requests or article viewing habits. In this respect, certain key words or terms, such as those relating to sports (i.e. “football” and “soccer”), can be detected within search requests and used to classify the user as someone interested in sports.” Culliss 3:13-36 “The present embodiment of the invention utilizes personal data to further refine search results . . . . Personal activity data includes data about past actions of the user, such as reading habits, viewing habits, searching habits, previous articles displayed or selected, previous search requests entered, previous or current site visits, previous key terms utilized within previous search results, and time or date of any previous activity.” Brookes 12:38-43 “creating and storing an interest profile for each database user indicative of categories of information of interest to said each database user, said interest profile comprising (i) a list of keywords taken from said finite hierarchical set and (ii) an associated priority level value for each keyword.” Brookes *See also*, 1:66-2:3.

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
		be some form of word frequencies such as TF/IDF and potentially other attributes such as the source, reading level, and length of the article.” Herz *See also* Abstract; 1:18-43; 4:–8:8; 55:44–56:14; 56:15-30; 58:57–60:9; Figures 1-16.	Chislenko 3:38-39 “Each user profile associates items with the ratings given to those items by the user. Each user profile may also store information in addition to the user’s ratings.” Chislenko 4:15-18 “For example, the system may assume that Web sites for which the user has created “bookmarks” are liked by that user and may use those sites as initial entries in the user’s profile.” Chislenko 4:40-50 “Ratings can be inferred by the system from the user’s usage pattern. For example, the system may monitor how long the user views a particular Web page and store in that user’s profile an indication that the user likes the page, assuming that the longer the user views the page, the more the user likes the page. Alternatively, a system may monitor the user’s actions to determine a rating of a particular item for the user. For example, the system may infer that a user likes an item which the user mails to many people and enter in the user’s profile and indication that the user likes that item.” Chislenko 21:64-22:2 “(a) storing, using the machine, a user profile in a memory for each of the plurality of users, wherein at least one of the user profiles includes a plurality of values, one of the plurality of values representing a rating given to one of a plurality of items by the user and another of the plurality of values representing additional information.”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			Chislenko 22:29-35 “storing, using the machine, a user profile in a memory for each of the plurality of users, wherein at least one of the user profiles includes a plurality of values, one of the plurality of values representing a rating given to one of a plurality of items by the user and another of the plurality of values representing information relating to the given ratings.” Dasan 3:21-24 “The present invention is a method and apparatus for automatically scanning information using a user-defined profile, and providing relevant stories from that information to a user based upon that profile.” Dasan 4:1-25 “[T]he user is able to connect to the remote server and specify a user profile, setting forth his interests. The user is able to specify the context for the information to be searched (e.g. the date). The user is able to save the profile on the remote machine. Finally the user is able to retrieve the personal profile (with any access control, if desired) and edit (add or delete entries) and save it for future operations. Dasan 4:34-39 “Using this interface, and HTTP, the server may notify the client of the results of that execution upon completion. The server’s application program, the personal newspaper generator maintains a record of the state of each user’s profile, and thus, provides state functionality from session to session to an otherwise stateless protocol.”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			Dasan *See, e.g.*, 5:37-6:3; 8:53-67. Dedrick 7:28-38 “Data is collected for personal profile database 27 by direct input from the end user and also by client activity monitor 24 monitoring the end user's activity. When the end user consumes a piece of electronic information, each variable (or a portion of each variable) within the header block for that piece of electronic information is added to the database for this end user. For example, if this piece of electronic information is made available to the end user for consumption in both audio and video format, and the end user selects the audio format, then this choice of format selection is stored in personal profile database Z1 for this end user.” Dedrick 3:54–4:4 “The GUI may also have hidden fields relating to "consumer variables." Consumer variables refer to demographic, psychographic and other profile information. Demographic information refers to the vital statistics of individuals, such as age, sex, income and marital status. Psychographic information refers to the lifestyle and behavioral characteristics of individuals, such as likes and dislikes, color preferences and personality traits that show consumer behavioral characteristics. Thus, the consumer variables refer to information such as marital status, color preferences, favorite sizes and shapes, preferred learning modes, employer, job title, mailing address, phone number, personal and business areas of interest, the willingness to participate in a survey, along with various lifestyle information. This

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			information will be referred to as user profile data, and is stored on a consumer owned portable profile device such as a Flash memory-based PCMClA pluggable card.” Dedrick *See, e.g.*, Abstract, Figures 1-8. Eichstaedt 1:34-43 “The present invention provides a profiling technique that generates user interest profiles by monitoring and analyzing a user’s access to a variety of hierarchical levels within a set of structured documents, e.g., documents available at a web site. Each information document has parts associated with it and the documents are classified into categories using a known taxonomy. In other words, each document is hierarchically structured into parts, and the set of documents is classified as well.” Eichstaedt 3:28-31 “The profile generation algorithm in the present embodiment learns from positive feedback. Each view of a document signifies an interest level in the content of the document.” Eichstaedt 1:43-55 “In other words, each document is hierarchically structured into parts, and the set of documents is classified as well. The user interest profiles are automatically generated based on the type of content viewed by the user. The type of content is determined by the text within the parts of the documents viewed and the classifications of the documents viewed. In addition, the profiles also are generated based

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			on other factors including the frequency and currency of visits to documents having a given classification, and/or the hierarchical depth of the levels or parts of the documents viewed. User profiles include an interest category code and an interest score to indicate a level of interest in a particular category. Unlike static registration information, the profiles in this invention are constantly changing to more accurately reflect the current interests of an individual.” Eichstaedt 2:15-41 “A preferred embodiment of the present invention automatically generates a profile that accurately captures a user’s stable interest after monitoring the user’s interaction with a set of structured documents. The technique of the present embodiment is based on the following three assumptions. First, each document in the corpus has different levels, parts, or views. These views are used to determine the level of interest a user has in a particular document. A hierarchical document structure is a good example for a document with different views. Structured documents such as patents have a title, an abstract and a detailed description. These parts of the document may be categorized according to a 3-level hierarchy which then can be used to determine how interested a user is in a particular topic. For example, if a user only views the title of a patent document, the user probably has little or no interest in the content of the document. If the user views the abstract as well, the user can be assumed to have more interest in the content of the document. If the user goes on to

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			view the detailed description, then there is good evidence that the user has a strong interest in the document, and the category into which it is classified. Generally, the more views, levels, or parts a document has, the finer will be the granularity of the present system. Although not all documents are structured at present, with the advent of XML, it is likely that the proportion of hierarchical documents available on the internet and in other databases will only increase." Eichstaedt 3:15-18 "In the system of the present invention, a special access analyzer and profile generator 62 analyzes information about user access to database 60 to generate a profile for the user. The profile is then used by a webcasting system 64 to provide or "push" customized information back to the user 54." Eichstaedt 5:32-36 "The automatic profile generation algorithm is completely automated and derives the user profiles from implicit feedback. Therefore, the user community does not have to learn new rules to customize the pushed information stream." Krishnan 2:37-41 "The information access monitor computes user/group profiles to identify information needs and interests within the organization and can then automatically associate users/groups with information of relevance." Krishnan 4:1-4 "[A] profile of a user's attributes is termed a 'user profile'; a summary of digital profiles of objects accessed by a user

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			and/or noted as of interest to the user, is termed the 'interest summary' of that user." Krishnan *See also* Fig. 6. Reese 4:35-53 "The user profile is intended to focus the retrieved results on meaningful data. One type of user profile is related to the demographics of the user. For example, the user profile might include the area code, zip code, state, sex, and age of a user. With such a profile, the matching server would retrieve data to the client related to the client's demographics. For example, if the user were interested in current events in the state of Oregon, the matching server would retrieve data and compile an aggregate database relating to current events pertinent to the user's age and area, e.g., Portland. Similarly, if the user sought information regarding retail purchases, the matching server would retrieve data relevant to the user's demographics. A demographics user profile is also very effective for advertisers that wish to advertise their goods or services on the matching server so that specific advertisements can be targeted at user's with specific user profile demographics. Other user profiles include, but are not limited to, areas of interest, business, politics, religion, education, etc." Reese 5:55-65 "The user profile form 600 includes a Search Type field 630 that allows a user to select whether the user wants an exact match of the user profile with the search data or whether the user will accept some lesser amount of exactness as acceptable for

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			retrieved data. The user profile form 600 further allows the user to enter demographics specific to the user. In FIG. 6, the demographics include area code 640, zip code 650, state 660, sex 670, age 680, and some other identifiers 690. Once the user enters the appropriate data in the user profile form 600, the user is instructed to save the profile by a "Save Profile" 694 button." Reese 8:26-35 "Thus far, the invention is focused on a user-created user profile. The invention also contemplates that the user profile may be constructed by the client based on the user's search habits. In other words, an artificial intelligence system may be created to develop a user profile. In the same way that a system is trained to be associative with regard to matching profile elements, the entire profile may be trained based on a user's search habits. For instance, a user profile that relates to demographics can be trained by recognizing user habits relating to demographics." Sheena 4:40-49 "Ratings can be inferred by the system from the user's usage pattern. For example, the system may monitor how long the user views a particular Web page and store in that user's profile an indication that the user likes the page, assuming that the longer the user views the page, the more the user likes the page. Alternatively, a system may monitor the user's actions to determine a rating of a particular item for the user. For example, the system may infer that a user likes an item which the user mails to many people and enter in the user's profile an indication that the user

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			likes that item.” Sheena 2:9-14 “In one aspect the present invention relates to a method for recommending an item to one of a plurality of users. The method begins by storing a user profile in a memory by writing user profile data to a memory management data object. Item profile data is also written to a memory management data object.” Sheena 3:34-67 “Each user profile associates items with the ratings given to those items by the user. Each user profile may also store information in addition to the user’s rating. In one embodiment, the user profile stores information about the user, e.g. name, address, or age. In another embodiment, the user profile stores information about the rating, such as the time and date the user entered the rating for the item. User profiles can be any data construct that facilitates these associations, such as an array, although it is preferred to provide user profiles as sparse vectors of n-tuples. Each n-tuple contains at least an identifier representing the rated item and an identifier representing the rating that the user gave to the item, and may include any number of additional pieces of information regarding the item, the rating, or both. Some of the additional pieces of information stored in a user profile may be calculated based on other information in the profile, for example, an average rating for a particular selection of items (e.g., heavy metal albums) may be calculated and stored in the user’s profile. In some embodiments, the profiles are provided

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			as ordered n-tuples. Alternatively, a user profile may be provided as an array of pointers; each pointer is associated with an item rated by the user and points to the rating and information associated with the rating. A profile for a user can be created and stored in a memory element when that user first begins rating items, although in multi-domain applications user profiles may be created for particular domains only when the user begins to explore, and rate items within, those domains. Alternatively, a user profile may be created for a user before the user rates any items in a domain. For example, a default user profile may be created for a domain which the user has not yet begun to explore based on the ratings the user has given to items in a domain that the user has already explored.” Sheena 28:16-21 “(a) storing a user profile, in the memory, for each of a plurality of users, wherein the user profile comprises a separate rating value, supplied by a particular one of the users, for each corresponding one of a plurality of items, said items including the item non-rated by the user.” Siefert 2:48-59 “In addition, in other forms of the invention, a profile is maintained which specifies certain preferences of the user. Two such preferences are (1) a preferred natural language (such as English or French), (2) the type of interface which the user prefers. The invention presents the resource in a manner compatible with the profile. Also, another profile, termed a “learning profile:’ is maintained, which, in a simplified sense,

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			specifies the current status of a user. with respect to a curriculum which the user is undertaking. The invention ensures compatibility between the resource and the learning profile, if possible." Siefert 8:60-62 "As stated above, the user profile contains information identifying the preferences of the user." Siefert 11:57-63 "The user profile specifies preferences of a user. It may not be possible, in all cases, to cause a resource selected by a user to become compatible with all specified preferences. However, insofar as the resource is transformed so that more preferences are matched than previously, the invention can be said to "enhance" the compatibility between the resource and the preferences." Belkin p. 397 "The search intermediary uses his knowledge about the IR system (with its data collections) and the searcher to formulate requests directly to the IR system. The search intermediary has formulated a model of the user and taken advantage of his existing model of the IR system." Belkin p. 399 "In the general information seeking interaction, the IR system needs to have (see Table 1 for a brief listing of the ten functions and their acronyms): a model of the user himself, including goals, intentions and experience (UM)." Han p. 409 "Personalized Web Agents Another group of Web agents includes those

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			that obtain or learn user preferences and discover Web information sources that correspond to these preferences, and possibly those of other individuals with similar interests (using collaborative filtering)" Han p. 409 "As the user browses the Web, the profile creation module builds a custom profile by recording documents of interest to the user. The number of times a user visits a document and the total amount of time a user spends viewing a document are just a few methods for determining user interest [1, 3, 4]. Once WebACE has recorded a sufficient number of interesting documents, each document is reduced to a document vector and the document vectors are passed to the clustering modules." Menczer p. 158-9 "Words are the principal asset in text collections, and virtually all information retrieval systems take advantage of words to describe and characterize documents, query, and concepts such as "relevance" or "aboutness" . . . This metric can be called word topology and is the reason why documents are usually represented as word vectors in information retrieval . . . [l]inks, constructed manually to point from one page to another, reflect an author's attempts to relate her writings to others.' Word topology is a epiphenomenal consequence of word vocabulary choices made by many authors, across many pages. The entire field of free text information retrieval is based on the statistical patterns reliably present in such vocabulary usage. By making our agents

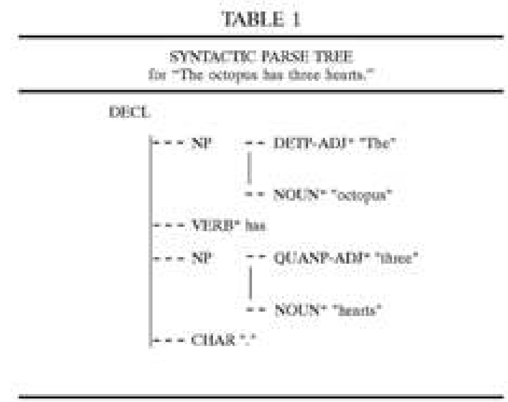
The '067 Patent	Salton '89	Herz	Additional Prior Art References
			perceptually sensitive to word topology features." Menczer p. 160 "For the reasons outlined in Section 2, each agent's genotype also contains a list of keywords, initialized with the query terms." [Agent's genotype is its version of a user profile.] Menczer p. 163 "The user initially provides a list of keywords and a list of starting points, in the form of a bookmark file." [The bookmarks and starting points are evidence of the profile the agent uses in creating its genotype.] Armstrong p. 1 "In interactive mode, WebWatcher acts as a learning apprentice [Mitchell et al., 1985; Mitchell et. al., 1994], providing interactive advice to the Mosaic user regarding which hyperlinks to follow next, then learning by observing the user's reaction to this advice as well as the eventual success or failure of the user's actions." Armstrong p. 4 "1. *Underlined words in the hyperlink.* 200 boolean features are allocated to encode selected words that occur within the scope of the hypertext link (i.e., the underlined words seen by the user). These 200 features correspond to only the 200 words found to be most informative over all links in the training data (see below.)" Armstrong p. 4: "The task of the learner is to learn the general function *UserChoice?*, given a sample of training data logged from users."

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
(b) constructing, by the remote computer system, a plurality of data item profiles, each plural data item profile corresponding to a different one of each plural data item stored in the remote data storage system, each of said plural data item profiles being representative of a second linguistic pattern of a corresponding plural data item, each said plural second linguistic pattern being substantially unique to each corresponding plural data item;	Salton ’89 p. 275. “[I]n these circumstances, it is advisable first to characterize record and query content by assigning special content descriptions, or profiles, identifying the items and representing text content. The text profiles can be used as short-form descriptions; they also serve as document, or query, surrogates during the text-search and [text]–retrieval operations.” Salton ’89 p. 294-6 (see also fn. 28-30) (*Linguistic methodologies including s*yntactic class indicators (adjective, noun, adverb, etc.) are assigned to the terms). Salton ’89 p. 389 (see also fn. 23-25) (Syntactic class markers, such as [noun], adjective, and pronoun, are first attached to the text words. Syntactic class patterns are then specified, such as “noun-noun”, or “adjective-adjective-noun,” and groups of text words corresponding to permissible syntactic class patterns are assigned to the texts for content identification. Word frequency and word distance constraints may also be used to refine phrase construction.” Salton ’89 p. 391, Fig. 11.3	Herz 79:11-22 “A method for cataloging a plurality of target objects that are stored on an electronic storage media, where users are connected via user terminals and bidirectional data communication connections to a target server that accesses said electronic storage media, said method comprising the steps of: storing on said electronic storage media each target object; automatically generating in said target server, target profiles for each of said target objects that are stored on said electronic storage media, each of said target profiles being generated from the contents of an associated one of said target objects and their associated target object characteristics” Herz 6:43-46 “The specific embodiment of this system disclosed herein illustrates the use of a first module which automatically constructs a “target profile” for each target object in the electronic media based on various descriptive attributes of the target object.”	Salton ‘68 p. 11 (Statistical association methods, Syntactic analysis methods, and Statistical phrase recognition methods). Salton ‘68 p. 30 “The word stem thesaurus and suffix list. One of the earliest ideas in automatic information retrieval was the suggested use of words contained in documents and search requests for purposes of content identification. No elaborate content analysis is then required, and the similarity between different items can be measured simply by the amount of overlap between the respective vocabularies.” Salton ‘68 p. 33 “The phrase dictionaries. Both the regular and the stem thesauruses are based on entries corresponding either to single words or to single word stems. In attempting to perform a subject analysis of written text, it is possible, however, to go further by trying to locate phrases consisting of sets of words that are judged to be important in a given subject area.” Salton ‘68 p. 35-36 “The syntactic phrase dictionary has a more complicated structure, as shown by the excerpt reproduced in Fig. 2-6. Here, each syntactic phrase, also known as criterion tree or criterion phrase, consists not only of a specification of the component concepts but also of syntactic indicators, as well as of syntactic relations that may obtain between the included concepts. . . . More specifically, there are four main classes of syntactic specifications, corresponding to noun phrases, subject-verb relations, verb-object

The '067 Patent	Salton '89	Herz	Additional Prior Art References
		Herz 12:54-13:53 "In particular, a textual attribute, such as the full text of a movie review, can be replaced by a collection of numeric attributes that represent scores to denote the presence and significance of the words "aardvark," "aback," "abacus," and so on through "zymurgy" in that text. The score of a word in a text may be defined in numerous ways. The simplest definition is that the score is the rate of the word in the text, which is computed by computing the number of times the word occurs in the text, and dividing this number by the total number of words in the text. This sort of score is often called the "term frequency" (TF) of the word. The definition of term frequency may optionally be modified to weight different portions of the text unequally: for example, any occurrence of a word in the text's title might be counted as a 3-fold or more generally k-fold occurrence (as if the title had been repeated k times within	relations, and subject-object relations." Braden 7:19-23 "Generally speaking and in accordance with our present invention, we have recognized that precision of a retrieval engine can be significantly enhanced by employing natural language processing to process, i.e., specifically filter and rank, the records, i.e., ultimately the documents, provided by a search engine used therein." Braden 11:62-14:61 "In general, to generate logical form triples for an illustrative input string, e.g. for input string 510, that string is first parsed into its constituent words. Thereafter, using a predefined record (not to be confused with document records employed by a search engine), in a stored lexicon, for each such word, the corresponding records for these constituent words, through predefined grammatical rules, are themselves combined into larger structures or analyses which are then, in turn, combined, again through predefined grammatical rules, to form even larger structures, such as a syntactic parse tree. A logical form graph is then built from the parse tree. Whether a particular rule will be applicable to a particular set of constituents is governed, in part, by presence or absence of certain corresponding attributes and their values in the word records. The logical form graph is then converted into a series of logical form triples. Illustratively, our invention uses such a lexicon having approximately 165,000 head word entries. This lexicon includes various classes of words, such as, e.g., prepositions, conjunctions, verbs, nouns,

The '067 Patent	Salton '89	Herz	Additional Prior Art References
		the text), in order to reflect a heuristic assumption that the words in the title are particularly important indicators of the text's content or topic. However, for lengthy textual attributes, such as the text of an entire document, the score of a word is typically defined to be not merely its term frequency, but its term frequency multiplied by the negated logarithm of the word's "global frequency," as measured with respect to the textual attribute in question. The global frequency of a word, which effectively measures the word's uninformativeness, is a fraction between 0 and 1, defined to be the fraction of all target objects for which the textual attribute in question contains this word. This adjusted score is often known in the art as TF/IDF ("term frequency times inverse document frequency"). When global frequency of a word is taken into account in this way, the common, uninformative words have scores comparatively close to zero, no matter how often or	operators and quantifiers that define syntactic and semantic properties inherent in the words in an input string so that a parse tree can be constructed therefor. Clearly, a logical form (or, for that matter, any other representation, such as logical form triples or logical form graph within a logical form, capable of portraying a semantic relationship) can be precomputed, while a corresponding document is being indexed, and stored, within, e.g., a record for that document, for subsequent access and use rather than being computed later once that document has been retrieved. Using such precomputation and storage, as occurs in another embodiment of our invention discussed in detail below in conjunction with FIGS. 10-13B, drastically and advantageously reduces the amount of natural language processing, and hence execution time associated therewith, required to handle any retrieved document in accordance with our invention. In particular, an input string, such as sentence 510 shown in FIG. 5A, is first morphologically analyzed, using the predefined record in the lexicon for each of its constituent words, to generate a so-called "stem" (or "base") form therefor. Stem forms are used in order to normalize differing word forms, e.g., verb tense and singular-plural noun variations, to a common morphological form for use by a parser. Once the stem forms are produced, the input string is syntactically analyzed by the parser, using the grammatical rules and attributes in the records of the constituent words, to yield the syntactic parse tree therefor. This tree depicts the structure of the input string, specifically each word or

<table>
<tr><th>The ‘067 Patent</th><th>Salton ’89</th><th>Herz</th><th>Additional Prior Art References</th></tr>
<tr><td></td><td></td><td>rarely they appear in the text. Thus, their rate has little influence on the object’s target profile. Alternative methods of calculating word scores include latent semantic indexing or probabilistic models. Instead of breaking the text into its component words, one could alternatively break the text into overlapping word bigrams (sequences of 2 adjacent words), or more generally, word n-grams. These word n-grams may be scored in the same way as individual words. Another possibility is to use character n-grams. For example, this sentence contains a sequence of overlapping character 5-grams which starts “for e”, “or ex”, “r exa’’, “exam”, “examp”, etc. The sentence may be characterized, imprecisely but usefully, by the score of each possible character 5-gram (“aaaaa”, “aaaab”, ... “zzzzz”) in the sentence. Conceptually speaking, in the character 5-gram case, the textual attribute would be decomposed into at least 265=11,881,376 numeric</td><td>phrase, e.g. noun phrase “The octopus”, in the input string, a category of its corresponding grammatical function, e.g., NP for noun phrase, and link(s) to each syntactically related word or phrase therein. For illustrative sentence 510, its associated syntactic parse tree would be:



A start node located in the upper-left hand corner of the tree defines the type of input string being parsed. Sentence types include “DECL” (as here) for a declarative sentence, “IMPR” for an imperative sentence and “QUES” for a question. Displayed vertically to the right and below the start node is a first level analysis. This analysis has a head node indicated by an asterisk, typically a main verb (here the word “has”), a premodifier (here the noun phrase “The octopus”), followed by a postmodifier (the noun phrase “three hearts”). Each leaf of the tree contains a lexical term or a punctuation mark. Here, as labels, “NP” designates a noun phrase, and “CHAR” denotes a punctuation mark. The syntactic</td></tr>
</table>

<table>
<tr><th>The ‘067 Patent</th><th>Salton ’89</th><th>Herz</th><th>Additional Prior Art References</th></tr>
<tr>
<td></td>
<td></td>
<td>attributes. Of course, for a given target object, most of these numeric attributes have values of 0, since most 5-grams do not appear in the target object attributes. These zero values need not be stored anywhere. For purposes of digital storage, the value of a textual attribute could be characterized by storing the set of character 5-grams that actually do appear in the text, together with the nonzero score of each one. Any 5-gram that is not included in the set can be assumed to have a score of zero. The decomposition of textual attributes is not limited to attributes whose values are expected to be long texts. A simple, one-term textual attribute can be replaced by a collection of numeric attributes in exactly the same way. Consider again the case where the target objects are movies. The “name of director” attribute, which is textual, can be replaced by numeric attributes giving the scores for “Federico-Fellini,” “Woody-Allen,” “Terence-Davies,” and so forth, in that</td>
<td>parse tree is then further processed using a different set of rules to yield a logical form graph, such as graph 515 for input string 510. The process of producing a logical form graph involves extracting underlying structure from syntactic analysis of the input string; the logical form graph includes those words that are defined as having a semantic relationship there between and the functional nature of the relationship. The “deep” cases or functional roles used to categorize different semantic relationships include:

TABLE 2
<table>
<tr><td>Dsub</td><td>deep subject</td></tr>
<tr><td>Dind</td><td>deep indirect object</td></tr>
<tr><td>Dobj</td><td>deep object</td></tr>
<tr><td>Dnom</td><td>deep predicate nominative</td></tr>
<tr><td>Dcmp</td><td>deep object complement.</td></tr>
</table>
To identify all the semantic relationships in an input string, each node in the syntactic parse tree for that string is examined. In addition to the above relationships, other semantic roles are used, e.g. as follows:

TABLE 3
<table>
<tr><td>PRED</td><td>predicate</td></tr>
<tr><td>PTCL</td><td>particle in two-part verbs</td></tr>
<tr><td>Ops</td><td>Operator, e.g. numerals</td></tr>
<tr><td>Nadj</td><td>adjective modifying a noun</td></tr>
<tr><td>Dadj</td><td>predicate adjective</td></tr>
<tr><td>PROPS</td><td>otherwise unspecified modifier that is a clause</td></tr>
<tr><td>MODS</td><td>otherwise unspecified modifier that is not a clause</td></tr>
</table>
Additional semantic labels are defined as well, for example:

TABLE 4
<table>
<tr><td>TmeAt</td><td>time at which</td></tr>
<tr><td>LocAt</td><td>location</td></tr>
</table>
To identify all the semantic relationships in an input string, each node in the syntactic parse tree for that string is examined. In addition to the above relationships, other semantic roles are used.</td>
</tr>
</table>

The '067 Patent	Salton '89	Herz	Additional Prior Art References
		attribute." Herz 79:11-23 "A method for cataloging a plurality of target objects that are stored on an electronic storage media, . . . said method comprising the steps of: . . . automatically generating in said target server, target profiles for each of said target objects that are stored on said electronic storage media, each of said target profiles being generated from the contents of an associated one of said target objects and their associated target object characteristics." Herz 5:7-11 "The system for electronic identification of desirable objects of the present invention automatically constructs both a target profile for each target object in the electronic media based, for example, on the frequency with which each word appears in an article relative to its overall frequency of use in all articles." Herz 10:63-67; 11:1-7 "However, a more sophisticated system would	In any event, the results of such analysis for input string 510 is logical form graph 515. Those words in the input string that exhibit a semantic relationship therebetween (such as, e.g. "Octopus" and "Have") are shown linked to each other with the relationship therebetween being specified as a linking attribute (e.g. Dsub). This graph, typified by graph 515 for input string 510, captures the structure of arguments and adjuncts for each input string. Among other things, logical form analysis maps function words, such as prepositions and articles, into features or structural relationships depicted in the graph. Logical form analysis also resolves anaphora, i.e., defining a correct antecedent relationship between, e.g., a pronoun and a co-referential noun phrase; and detects and depicts proper functional relationships for ellipsis. Additional processing may well occur during logical form analysis in an attempt to cope with ambiguity and/or other linguistic idiosyncrasies. Corresponding logical form triples are then simply read in a conventional manner from the logical form graph and stored as a set. Each triple contains two node words as depicted in the graph linked by a semantic relationship therebetween. For illustrative input string 510, logical form triples 525 result from processing graph 515. Here, logical form triples 525 contain three individual triples that collectively convey the semantic information inherent in input string 510. Similarly, as shown in FIGS. 5B-5D, for input strings 530, 550 and 570, specifically exemplary sentences "The octopus has three hearts and two lungs.", "The octopus has three hearts and it can

The '067 Patent	Salton '89	Herz	Additional Prior Art References
		consider a longer target profile, including numeric and associative attributes: (a.) full text of document . . . (d.) language in which document is written . . . (g.) length in words . . . (h.) reading level." Herz *See also* Abstract; 1:18-43; 4:49–8:8; 9:1–16:62; 26:43–27:43; 55:44–56:14; 56:52–57:10.	swim.", and "I like shark fin soup bowls.", logical form graphs 535, 555 and 575, as well as logical form triples 540, 560 and 580, respectively result. There are three logical form constructions for which additional natural language processing is required to correctly yield all the logical form triples, apart from the conventional manner, including a conventional "graph walk", in which logical form triples are created from the logical form graph. In the case of coordination, as in exemplary sentence "The octopus has three hearts and two lungs", i.e. input string 530, a logical form triple is created for a word, its semantic relation, and each of the values of the coordinated constituent. According to a "special" graph walk, we find in FIG. 540 two logical form triples "haveDobj- heart" and "have-Dobj-lung". Using only a conventional graph walk, we would have obtained only one logical form triple "have-Dobj-and". Similarly, in the case of a constituent which has referents (Refs), as in exemplary sentence "The octopus has three hearts and it can swim", i.e. input string 550, we create a logical form triple for a word, its semantic relation, and each of the values of the Refs attribute, in additional to the triples generated by the conventional graph walk. According to this special graph walk, we find in triples 560 the logical form triple "swim-Dsuboctopus" in addition to the conventional logical form triple "swim-Dsub-it". Finally, in the case of a constituent with noun modifiers, as in the exemplary sentence "I like shark fin soup bowls", i.e. input string 570, additional logical form triples are created to represent possible

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			internal structure of the noun compounds. The conventional graph walk created the logical form triples “bowl-Mods-shark”, “bowl-Modsfin” and “bowl-Mods-soup”, reflecting the possible internal structure [[shark] [fin] [soup] bowl]. In the special graph walk, we create additional logical form triples to reflect the following possible internal structures [[shark fin] [soup] bowl] and [[shark] [fin soup] bowl] and [[shark [fin] soup] bowl], respectively: “fin-Mods-shark”, “soup-Mods-fin”, and “soup-Mods-shark”. Inasmuch as the specific details of the morphological, syntactic, and logical form processing are not relevant to the present invention, we will omit any further details thereof. However, for further details in this regard, the reader is referred to co-pending United States patent applications entitled “Method and System for Computing Semantic Logical Forms from Syntax Trees”, filed Jun. 28, 1996 and assigned Ser. No. 08/674,610 and particularly “Information Retrieval Utilizing Semantic Representation of Text”, filed Mar. 7, 1997 and assigned Ser. No. 08/886,814; both of which have been assigned to the present assignee hereof and are incorporated by reference herein.” Braden 7:47-53 “each of the documents in the set is subjected to natural language processing, specifically morphological, syntactic and logical form, to produce logical forms for each sentence in that document. Each such logical form for a sentence encodes semantic relationships, particularly argument and adjunct structure, between words in a

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			linguistic phrase in that sentence.” Culliss 2:33-37 “The articles can each be associated with one or more of these key terms by any conceivable method of association now known or later developed. A key term score is associated with each article for each of the key terms. Optionally, a key term total score can also be associated with the article.” Ahn 2:32-34 “Also, a document tree and a document index table is maintained for each document (such as Document Dl).” Brookes 12:27-37 “storing in association with each information item in the database system a plurality of parameters including (i) at least one keyword indicative of the subject matter of said information item, and (ii) a priority level value for each information item, wherein said priority level value is selected from a predetermined set ·of priority level values, and wherein said at least one keyword is selected from a finite hierarchical set of keywords having a tree structure relating broad keywords to progressively narrower keywords.” Brookes *See also*, 1:57-65. Dedrick 15:41-44 “The metering server 14 is capable of storing units of information relating to the content databases of the publisher/advertiser, including the entire content database.”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			Dedrick See, e.g., Abstract, Figures 1-8. Eichstaedt 2:42-50 “The second assumption is that the documents must already be assigned to at least one category of a known taxonomy tree for the database. Notice, however, that this system works with any existing taxonomy tree and does not require any changes to a legacy system. FIG. 1 illustrates a taxonomy tree with six leaf categories 50. Each leaf category has an interest value associated with it. Taxonomies are available for almost all domain-specific document repositories because they add significant value for the human user.” Eichstaedt 1:34-43 “The present invention provides a profiling technique that generates user interest profiles by monitoring and analyzing a user’s access to a variety of hierarchical levels within a set of structured documents, e.g., documents available at a web site. Each information document has parts associated with it and the documents are classified into categories using a known taxonomy. In other words, each document is hierarchically structured into parts, and the set of documents is classified as well.” Krishnan 3:64-4:1 “[I]nformation, which is typically electronic in nature and available for access by a user via the Internet, is termed an ‘object’; a digitally represented profile indicating an object’s attributes is termed an ‘object profile.’” Krishnan 7:13-42 “The basic [document]

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			indexing operation comprises three steps, noted above as: filtering, word breaking, and normalization . . . . Once the content filter has operated on the source file, the word breaker step is activated to divide the received text stream from the content filter into words and phrases. Thus, the word breaker accepts a stream of characters as an input and outputs words . . . . The final step of indexing is the normalization process, which removes ‘noise’ words and eliminates capitalization, punctuation, and the like.” Krishnan *See also* Fig. 6. Kupiec 13:13-20 “In step 250 the match sentences retained for further processing in step 245 are analyzed to detect phrases they contain. The match sentences are analyzed in substantially the same manner as the input string is analyzed in step 220 above. The detected phrases typically comprise noun phrases and can further comprise title phrases or other kinds of phrases. The phrases detected in the match sentences are called preliminary hypotheses.” Reese 7:1-24 “In collecting the information that matches the query request, the server may collect different forms of information. First, the server may collect entire content site data, for example, entire files or documents on a particular content server. Instead, the server may collect key words from particular sites (e.g., files) on individual content servers, monitor how often such key words are used in a document, and construct a database based on

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			these key words (step 822). Another way of collecting data is through the collection of content summaries (step 824). In this manner, rather than entire files or documents being transmitted to the server and ultimately to the client, only summaries of the documents or files are collected and presented. The summaries offer a better description of the content of the particular files or documents than the key words, because the user can form a better opinion of what is contained in the abbreviated document or file based on summaries rather than a few key words. The summaries may be as simple as collective abstracts or may involve the matching server identifying often used key words and extracting phrases or sentences using these key words from the document. Finally, the invention contemplates that titles may also be retrieved by the matching server and submitted to the client rather than entire documents or files.” Sheena 2:14-15 “Similarity factors are calculated for each of the users and the similarity factors are used to select a neighboring user set for each user of the system.” Sheena 4:56-5:17 “Profiles for each item that has been rated by at least one user may also be stored in memory. Each item profile records how particular users have rated this particular item. Any data construct that associates ratings given to the item with the user assigning the rating can be used. It is preferred is to provide item profiles as a sparse vector of

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			n-tuples. Each n-tuple contains at least an identifier representing a particular user and an identifier representing the rating that user gave to the item, and it may contain other information, as described above in connection with user profiles. As with user profiles, item profiles may also be stored as an array of pointers. Item profiles may be created when the first rating” Siefert 8:22-33 “In a very simple sense, the expert identifies the language of a sample of words, by reading the sample. Then, the invention analyzes samples of each language, in order to find unique character- and word patterns (or other patterns). Now the invention can associate unique patterns with each language. The invention stores the unique patterns, together with the corresponding language identities, in a reference table. Later, to identify a language, the invention looks for the unique patterns within a sample of the language, such as in a file whose language is to be identified. When a pattern is found, the invention identifies the language containing it, based on the table.” Armstrong p. 4 “1. *Underlined words in the hyperlink.* 200 boolean features are allocated to encode selected words that occur within the scope of the hypertext link (i.e., the underlined words seen by the user). These 200 features correspond to only the 200 words found to be most informative over all links in the training data (see below.)”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
(c) providing, by the user to the local computer system, search request data representative of the user's expressed desire to locate data substantially pertaining to said search request data;	Salton ’89 p. 160 “Several types of query specifications can be distinguished. A simple query is one containing the value of a single search key. A range query contains a range of values for a single key – for example, a request for all the records of employee ages 22 to 25. A functional query is specified by using a function for the values for certain search keys, for example the age of employees exceeding a given stated threshold.”	Herz 66:52-61 “However, in a variation, the user optionally provides a query consisting of textual and/or other attributes, from which query the system constructs a profile in the manner described herein, optionally altering textual attributes as described herein before decomposing them into numeric attributes. Query profiles are similar to the search profiles in a user’s search profile set, except that their attributes are explicitly specified by a user, most often for one-time usage, and unlike search profiles, they are not automatically updated to reflect changing interests.” Herz See also Abstract; 1:18-43; 4:49-8::8; 55:44–5:14; 56:15-30; 58:57–60:9; Figures 1-16.	Salton ‘68 p. 7 “When the search criteria are based in one way or another on the contents of a document, it becomes necessary to use some system of content identification, such as an existing subject classification or a set of content identifiers attached to each item, which may help in restricting the search to items within a certain subject area and in distinguishing items likely to be pertinent from others to be rejected.” Salton ‘68 p. 413 “The user participates in the system by furnishing information about his needs and interests, by directing the search and retrieval operations accordance with his special requirements, by introducing comments out systems operations, by specifying output format requirements, and nearly by influencing file establishment and file maintenance procedures.” Braden 7:35-38 “Specifically, in operation, a user supplies a search query to system 5. The query should be in full-text (commonly referred to as “literal”) form in order to take full advantage of its semantic content through natural language processing.” Culliss 2:39-41 “[T]he invention can accept a search query from a user and a search engine will identify matched articles.” Culliss 12:41-51 “A method of organizing a plurality of articles comprising . . . (b) accepting a first search query from a first user having first personal data.”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			Ahn 3:37-42 “In step 408, the invention receives a user search request containing a keyword and determines whether the search request is directed to searching an individual document or a group of documents. If the search request is directed to searching an individual document, then step 414 is performed.” Brookes 8:48-54 “In this manner the information in the system may be augmented by input from the users, questions may be asked of specific users and responses directed accordingly. A collection of information items related in this manner is termed a ‘discussion’. The context of a discussion is defined by the parameters (especially keywords) of its constituent information items.” Brookes *See, e.g.*, 12:27-37 “storing in association with each information item in the database system a plurality of parameters including (i) at least one keyword indicative of the subject matter of said information item, and (ii) a priority level value for each information item, wherein said priority level value is selected from a predetermined set of priority level values, and wherein said at least one keyword is selected from a finite hierarchical set of keywords having a tree structure relating broad keywords to progressively narrower keywords.” Dasan 7:28-38 “the user specifies search terms used in the full-text search. These are illustrated in field 804. Any number of search terms may be used and the “l” character is

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			treated as a disjunction (“or”). Then. by selecting either of user interface objects 806 or 808, the user specifies whether the search terms are case sensitive or not. This is detected at step 706. At step 708, using either a scrollable list containing selectable item(s), as illustrated in field 810, or other means, the user specifies the search context(s) (the publications, newsfeeds, etc...) in which to search. By the selection of icon 812 or other commit means.” Dedrick *See, e.g.*, Figures 1-8, 8:20–9:24, 14:55–64. Krishnan 7:61-63 “The query screen allows a user to express a query by simply filling out fields in a form.” Krishnan 12:36-47 “[A] method for enhancing efficiencies with which objects retrieved from the Internet are maintained for access by the multiple members, the method comprising: . . . receiving a member-generated query for one or more objects that can be obtained from the Internet.” Krishnan *See also* Fig. 6. Kupiec 4:7-8 “The method begins by accepting as input the user’s question and a set of documents that are assumed to contain the answer to the question.” Reese 7:1-23 “In collecting the information that matches the query request, the server may collect different forms of information.”

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			Menczer p. 162 "Consider for example the following query: "Political institutions: The structure, branches and offices of government." Menczer p. 163 "The user initially provides a list of keywords and a list of starting points, in the form of a bookmark file.[2] In step (0), the population is initialized by pre-fetching the starting documents. Each agent is "positioned" at one of these document and given a random behavior (depending on the representation) and an initial reservoir of "energy". In step (2), each agent "senses" its local neighborhood by analyzing the text of the document where it is currently situated. This way, the relevance of all neighboring documents -those pointed to by the hyperlinks in the current document- is estimated. Based on these link relevance estimates, an agent "moves" by choosing and following one of the links from the current document." Armstrong p. 4 "4. *Words used to define the user goal*. These features indicate words entered by the user while defining the information search goal. In our experiments, the only goals considered were searches for technical papers, for which the user could optionally enter the title, author, organization, etc. (see Figure 3). All words entered in this way throughout the training set were included (approximately 30 words, though the exact number varied with the training set used in the particular experiment). The encoding of the boolean feature in this case is assigned a 1 if

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			and only if the word occurs in the user-specified goal and occurs in the hyperlink, sentence, or headings associated with this example."
(d) extracting, by one of the local computer system and the remote computer system, a search request profile from said search request data, said search request profile being representative of a third linguistic pattern of said search request data;	Salton '89 p. 275 "In these circumstances, it is advisable first to characterize record and query content by assigning special content descriptions, or profiles, identifying the items and representing text content. The text profiles can be used as short-form descriptions; they also serve as document, or query, surrogates during the text-search and [text]–retrieval operations." Salton '89 p. 294-6 (see also fn. 28-30) (*Linguistic methodologies including syntactic* class indicators (adjective, noun, adverb, etc.) are assigned to the terms).	Herz 66:52-61 "However, in a variation, the user optionally provides a query consisting of textual and/or other attributes, from which query the system constructs a profile in the manner described herein, optionally altering textual attributes as described herein before decomposing them into numeric attributes. Query profiles are similar to the search profiles in a user's search profile set, except that their attributes are explicitly specified by a user, most often for one-time usage, and unlike search profiles, they are not automatically updated to reflect changing interests." Herz See also Abstract; 1:18-43; 4:49-8:8; 55:44–5:14; 56:15-30; 58:57–60:9; Figures 1-16.	Salton '68 p. 7 "In most of the semimechanized centers where the search operation is conducted automatically, it is customary to assign to documents and search requests alike a set of content identifiers, normally chosen from a controlled list of allowable terms, and to compare their respective lists of content identifiers in order to determine the similarity between stored items and requests for information. A simplified chart of the search and retrieval operations is shown in Fig. 1-2." Salton '68 p. 11 (Statistical association methods, Syntactic analysis methods, and Statistical phrase recognition methods). Salton '68 p. 30 "The word stem thesaurus and suffix list. One of the earliest ideas in automatic information retrieval was the suggested use of words contained in documents and search requests for purposes of content identification. No elaborate content analysis is then required, and the similarity between different items can be measured simply by the amount of overlap between the respective vocabularies." Salton '68 p. 33 "The phrase dictionaries. Both the regular and the stem thesauruses are based on entries corresponding either to single

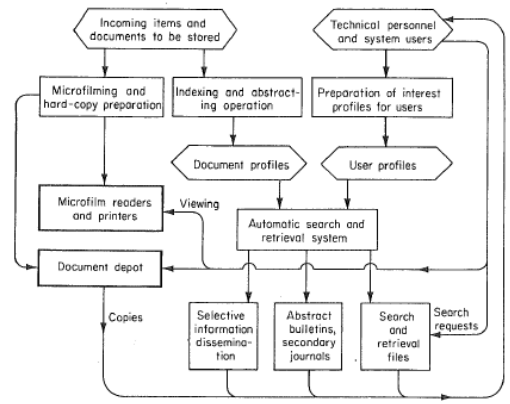
The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			words or to single word stems. In attempting to perform a subject analysis of written text, it is possible, however, to go further by trying to locate phrases consisting of sets of words that are judged to be important in a given subject area.” Salton ‘68 p. 34 “The statistical phrase dictionary is based on a phrase detection algorithm which takes into account only the statistical co-occurrence characteristics of the phrase components; specifically a statistical phrase is recognized if and only if all phrase components are present within a given document or within a given sentence of a document, and no attempt is made to detect any particular syntactic relation between the components. On the other hand, the syntactic phrase dictionary includes not only the specification of the particular phrase components that are to be detected but also information about the permissible syntactic dependency relations that must obtain if the phrase is to be recognized.” Salton ‘68 p. 35-36 “The syntactic phrase dictionary has a more complicated structure, as shown by the excerpt reproduced in Fig. 2-6. Here, each syntactic phrase, also known as criterion tree or criterion phrase, consists not only of a specification of the component concepts but also of syntactic indicators, as well as of syntactic relations that may obtain between the included concepts. . . . More specifically, there are four main classes of syntactic specifications, corresponding to noun phrases, subject-verb relations, verb-object

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			relations, and subject-object relations.” Braden 7:19-23 “Generally speaking and in accordance with our present invention, we have recognized that precision of a retrieval engine can be significantly enhanced by employing natural language processing to process, i.e., specifically filter and rank, the records, i.e., ultimately the documents, provided by a search engine used therein.” Braden 11:1-4 “In addition, though not specifically shown, process 600 also internally analyzes the query to produce its corresponding logical form triples which are then locally stored within computer 300.” Braden *See, e.g.*, 11:62-14:61. Culliss 8:40-45 “One way to determine which personal data characteristics result in different query rankings is to compare the previous user relevancy scores, or ranking determined at least in part by the previous user relevancy scores, of queries, key terms or key term groupings in which a particular personal data characteristic is different.” Culliss 7:15-18 “Another embodiment of the present invention keeps track of the full queries, or portions thereof such as key terms groupings, which are entered by users having certain personal data characteristics. In this embodiment, queries or portions thereof such as key term groupings, are stored within an index, preferably along with the personal data and a previous-user relevancy score for each

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			query.” Dedrick *See, e.g.*, Figures 1-8, 8:20–9:24, 14:55–64. Krishnan 7:52-54 “The document search engine DSE converts Internet queries into a query form that is compatible with document search engine DSE indexes.” Krishnan 8:28-30 “The user at step 601 generates a query on the user’s client processor, such as client processor C1, as described above.” Krishnan *See also* Fig. 6. Kupiec 3:23-29 “The present invention provides a method for answer extraction. A system operating according to this method accepts a natural-language input string such as a user supplied question and a set of relevant documents that are assumed to contain the answer to the question. In response, it generates answer hypotheses and finds these hypotheses within the documents.” Kupiec 4:13-18 “The method then analyzes the question to detect the noun phrases that it contains. In this example, the noun phrases are “Pulitzer Prize,” “novelist,” “mayor,” and “New York City.” The method assumes that the documents contain some or all these noun phrases. This will be the case if the IR queries used to retrieve the primary documents are constructed based on the noun phrases.”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			Kupiec 11:33-12:46 “In step 310 noun phrases are detected. A noun phrase is a word sequences that consists of a noun, its modifiers such as adjectives and other nouns, and possibly a definite or indefinite article. . . . In step 315 main verbs are detected. Main verbs are any words that are tagged in step 300 as verbs and that are not auxiliary verbs. Typically there is one main verb in the input string, but there can also be none, or two or more. . . . In step 330 the results of steps 310, 315, and 320 are stored. The stored results represent the completed analysis of the input string. The results can be stored, for example, in a list of 3-tuples, one 3-tuple for each noun phrase, main verb, and title phrase detected during steps 310, 315, and 320. Each 3-tuple is an ordered list of the form (i, phrase-type, 25 text), where i is a unique index number associated with the phrase, such as its position (first, second, third ...) in the list; phrase-type indicates the type of phrase (noun phrase, main verb, or title phrase); and text is a string that contains the text of the phrase itself. . . in some embodiments an empty list is created as part of step 330 at the outset, prior to the execution of steps 310, 315, and 320, and thereafter is filled in incrementally during the processing of the steps 310, 315, and 320, so that upon completion of steps 310, 315, and 320, step 330 is effectively completed as well.” Han p.413: “The characteristic words of a cluster of documents are the ones that have document frequency and high average text frequency. . . . We define the TF word list as

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			the list of *k* words that have the highest average text frequency and the DF word list as the list of *k* words that have the highest document frequency. . . The query can be formed as $(c_1 \wedge c_2 \ldots \wedge c_m) \wedge (t_1 \vee t_2 \ldots \vee t_n)$ where $c_1 = TF \cap DF$ and $t_1 = TF - DF$." Menczer p. 162 "After noise words have been removed and the remaining words have been stemmed, the query is reduced to POLIT, INSTITUT, STRUCTUR BRANCH OFFIC GOVERN." Armstrong p. 4 "4. *Words used to define the user goal*. These features indicate words entered by the user while defining the information search goal. In our experiments, the only goals considered were searches for technical papers, for which the user could optionally enter the title, author, organization, etc. (see Figure 3). All words entered in this way throughout the training set were included (approximately 30 words, though the exact number varied with the training set used in the particular experiment). The encoding of the boolean feature in this case is assigned a 1 if and only if the word occurs in the user-specified goal and occurs in the hyperlink, sentence, or headings associated with this example."

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
(e) determining, by one of the local computer system and the remote computer system, a first similarity factor representative of a first correlation between said search request profile and said user profile by comparing said search request profile to said user profile;	Salton ’89 p. 317-9 “As a matter of practice, the vector-space model can then be used to obtain correlations, or similarities, between pairs of stored documents, or between queries and documents, under the assumption that the *t* term vectors are orthogonal, or that the term vectors are linearly independent, so that a proper basis exists for the vector space. When term dependencies or associations are available from outside sources, they can be taken into account . . . A list of typical vector-similarity measures appears in table 10.1 . . . Table 10.1 Measures of vector similarity. Cosine coefficient $$\frac{\sum_{i=1}^{t} x_i \bullet y_i}{\sqrt{\sum_{i=1}^{t} x_i^2 \bullet \sum_{i=1}^{t} y_i^2}}$$ . . . Some of the advantages are the model’s simplicity, the ease with which it accommodates weighted terms, and its provision of ranked retrieval output in decreasing order of query-document similarity.”	Herz 14:40-15:13 “Similarity Measures. What does it mean for two target objects to be similar? More precisely, how should one measure the degree of similarity? Many approaches are possible and any reasonable metric that can be computed over the set of target object profiles can be used, where target objects are considered to be similar if the distance between their profiles is small according to this metric. Thus, the following preferred embodiment of a target object similarity measurement system has many variations. First, define the distance between two values of a given attribute according to whether the attribute is a numeric, associative, or textual attribute. If the attribute is numeric, then the distance between two values of the attribute is the absolute value of the difference between the two values. (Other definitions are also possible: for example, the distance between prices pl and p2 might be defined by l (Plp2)	Salton ‘68 p. 414, Fig. 10-4.  **Fig. 10-4** Typical technical information center. Braden 11:22-26 “Thereafter, through comparing the logical form triples for the query against those for each document, process 600 scores each document that contains at least one matching logical form triple, then ranks these particular documents based on their scores.” Braden 17:44-53 “Of these triples, two are identical, i.e., “HAVE-Dsub-OCTOPUS”. A score for a document is illustratively a numeric sum of the weights of all uniquely matching triples in that document. All duplicate matching triples for any document are ignored. An illustrative ranking of the relative weightings of the different types of relations that can occur in a triple, in descending order from their largest to smallest weightings are: first, verb-object combinations (Dobj); verb-subject combinations (Dsub); prepositions and operators (e.g. Ops), and finally modifiers (e.g. Nadj).”

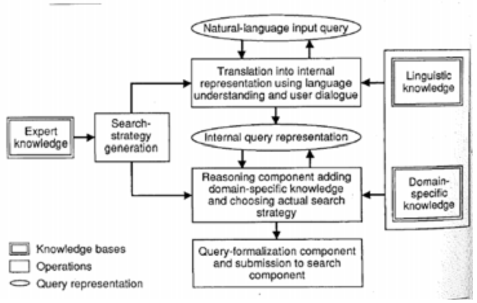
The '067 Patent	Salton '89	Herz	Additional Prior Art References
		1/(max(pl,p2)+I), to recognize that when it comes to customer interest, $5000 and $5020 are very similar, whereas $3 and $23 are not.) If the attribute is associative, then its value V may be decomposed as described above into a collection of real numbers, representing the association scores between the target object in question and various ancillary objects. V may therefore be regarded as a vector with components V1, V2, V3 etc., representing the association scores between the object and ancillary objects 1, 2, 3, etc., respectively. The distance between two vector values V and U of an associative attribute is then computed using the angle distance measure, arccos (VU'/sqrt((Vv')(UU')). (Note that the three inner products in this expression have the form XY'=X1 Y1+X2 Y2+X3 Y3+... , and that for efficient computation, terms of the form Xi Y, may be omitted from this sum if either of the scores Xi and Y, is zero.) Finally, if the attribute is	Braden 25:41-48 "Rather than using fixed weights for each different attribute in a logical form triple, these weights can dynamically vary and, in fact, can be made adaptive. To accomplish this, a learning mechanism, such as, e.g., a Bayesian or neural network, could be appropriately incorporated into our inventive process to vary the numeric weight for each different logical form triple to an optimal value based upon learned experiences." Culliss 10:47-52 "To present personalized search results to a particular person searching with a particular term or query, the present invention may display a number of articles from a number of the narrower related key term groupings or queries which are ranked by their respective previous-user relevancy scores." Culliss 11:11-20 "It is also possible to consider both the previous-user relevancy score of the top narrower related key term groupings or queries, as well as the previous-user relevancy score of the articles under these narrower related key term groupings or queries. In this respect, the previous-user relevancy score of the top narrower related key term groupings or queries and the previous-user relevancy score of the articles under these narrower related key term groupings or queries can be combined in any possible manner, such as by adding, multiplying, or averaging together."

The '067 Patent	Salton '89	Herz	Additional Prior Art References
		textual, then its value V may be decomposed as described above into a collection of real numbers, representing the scores of various word n-grams or character n-grams in the text. Then the value V may again be regarded as a vector, and the distance between two values is again defined via the angle distance measure. Other similarity metrics between two vectors, such as the dice measure, may be used instead." Herz 1:25-28; 4:55-62 Herz contemplates using both "user profiles" and "query profiles" to form "target profile interest summaries" that "describe[] the user's interest level in various types of target objects." Herz 56:19-28 Herz further teaches that search profiles can be determined by "asking the user to specify search profiles directly by giving keywords and/or numeric attributes" (the search request/query profile) *and* by "using copies of the profiles of target objects or target clusters that the user	Culliss 5:18-21 "When a user first enters a search query, the personal data can be considered part of the request and stored within or added to the index, individually or in groupings with other items of data such as key terms, categories, or ratings." Culliss 5:41-45 "When the next user enters a search request, the search request and the user's personal data are combined to form groupings containing key term groupings, key terms and personal data groupings, category and personal data groupings, rating and personal data groupings, etc." Culliss 10:8-13 "For example, when a woman enters the search request 'shoes,' the system can look for narrower related queries or key term groupings which contain or are related to the term 'shoes' and which have been entered by previous users having similar personal data, such as that of being a 'woman.'" Dedrick See, e.g., Figures 1-8, 8:20–9:24, 14:55–64. Krishnan 8:34-45 "The information access monitor IAM, at step 604, uses the relevance index information stored in the index files IF to process the request and identify the ones of the objects previously indexed by document search engine DSE which match the relevance index information stored in index files IF. This is accomplished by performing an object relevance determination based upon the identity of the user requesting the information, the user's profile and user's interest summary

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
		indicates are representative of his or her interest” (the user profile). Herz 57:23-27 *Both* types of data are to be considered in determining which documents are most likely of interest to the user.	indexes stored in the database DB, and other user profile criteria, administrative criteria, and object characterizing data.” Krishnan *See also* Fig. 6. Kupiec 18:1-26 “6.5 Matching Templates Against Primary Documents. In step 264 an attempt is made to verify the linguistic relation under consideration for the hypothesis under consideration in the context of the primary documents. This is done by matching the filled-in templates generated in step 263 against the primary documents. In other words, sentences in which the hypothesis appears in the context of a template are sought in the primary documents. Any such sentences found are retained in association with the hypothesis as verification evidence for use in later processing steps. For example, if the template is “NP(Justice) (is, was) X” and the hypothesis is “Earl Warren,” the filled-in template is “NP(Justice) (is, was) Earl Warren,” and documents containing sentences such as “At that time the Chief Justice was Earl Warren . . . “ are potential matches. As another example, if the template is “X succeeded Shastri” and the hypothesis is “Indira Gandhi,” the filled-in template is “Indira Gandhi succeeded Shastri.” The answer extraction subsystem seeks one or more primary documents that contain sentences conforming to this filled-in template, for example, “Indira Gandhi succeeded Shastri .... “ The testing of step 264 is carried out using only the primary documents. If sufficient template matches are

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			found among the primary documents, then the linguistic relation is considered verified. In this case it is unnecessary to run secondary queries and steps 265 and 266 are skipped for this linguistic relation and hypothesis." Reese 3:45-55 "The invention contemplates that the matching server 120 works with the client user profile request 100 to pare down the data delivered to the client. The matching server 120 pre-selects an aggregate of data that is determined to be the most relevant to different sets of user profile requests 100. The matching server 120 does this by searching various content sites 130, 140, 150, 160 on the Internet or other network. A user profile request 100 is applied against the matching server 120 aggregate of data like a sieve, and only data matching the user profile request 100 is returned to the client 110." Belkin p. 396 "As online search systems tend to rely on specialized access mechanisms--commands. index terms, query forms--it is natural to seek effective, automatic ways of mapping the user's request onto a search query, both because assistance by human intermediaries is costly and because it would be nice to offer the end-user direct access to the search system, . . . there is also the important business of establishing the user's real need, so a more significant function of an intelligent interface could be to help the user explicitly formulate a statement of his need." Menczer p. 162 "This is all the initial population knows about what the user is

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			interested in. But after some of the visited documents are assessed by the user, her preferences become better defined . . . This list captures an image of what word features are best correlated with relevance. The term COURT, for example, appears to have the highest correlation with relevance even though it was not a part of the query.” Armstrong p. 4 “In each case, the words were selected by first gathering every distinct word that occurred over the training set, then ranking these according to their mutual information with respect to correctly classifying the training data.”
(f) determining, by one of the local computer system and the remote computer system, a plurality of second similarity factors, each said plural second similarity factor being representative of a second correlation between said search request profile and a different one of said plural data item profiles, by comparing said search request profile to each of said plural data item profiles;	Salton ’89 p. 306 A similarity factor is represented by the following equation: $$sim(Q, D_i) = \frac{\sum_{j=1}^{t} w_{qj} \bullet d_{ij}}{\sqrt{\sum_{j=1}^{t} (d_{ij})^2 \bullet \sum_{j=1}^{t} (w_{qj})^2}}$$ where: Q = query; D = document; W_{qi} = inverse document-frequency weights D_{ij} = term-frequency and inverse document-frequency weights. Salton ’89 p. 366 “Figure 10.20 Expert interface system for text retrieval. [73]”	Herz 14:40-15:13 “Similarity Measures. What does it mean for two target objects to be similar? More precisely, how should one measure the degree of similarity? Many approaches are possible and any reasonable metric that can be computed over the set of target object profiles can be used, where target objects are considered to be similar if the distance between their profiles is small according to this metric. Thus, the following preferred embodiment of a target object similarity measurement system has	Salton ‘68 p. 11 7. “Request-document matching procedures which make it possible to use a variety of different correlation methods to compare analyzed documents with analyzed requests, including concept weight adjustments and variations in the length of the document texts being analyzed.” Salton ‘68 p. 414, Fig. 10-4. Braden 11:22-26 “Thereafter, through comparing the logical form triples for the query against those for each document, process 600 scores each document that contains at least one matching logical form triple, then ranks these particular documents based on their scores.” Braden 17:44-53 “Of these triples, two are

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
	Figure 10.20 Expert interface system for text retrieval [73].  Salton ’89 p. 317-319 “As a matter of practice, the vector-space model can then be used to obtain correlations, or similarities, between pairs of stored documents, or between queries and documents, under the assumption that the *t* term vectors are orthogonal, or that the term vectors are linearly independent, so that a proper basis exists for the vector space. When term dependencies or associations are available from outside sources, they can be taken into account . . . A list of typical vector-similarity measures appears in table 10.1 . . . Table 10.1 Measures of vector similarity.” Cosine coefficient $$\frac{\sum_{i=1}^{t} x_i \bullet y_i}{\sqrt{\sum_{i=1}^{t} x_i^2 \bullet \sum_{i=1}^{t} y_i^2}}$$	many variations. First, define the distance between two values of a given attribute according to whether the attribute is a numeric, associative, or textual attribute. If the attribute is numeric, then the distance between two values of the attribute is the absolute value of the difference between the two values. (Other definitions are also possible: for example, the distance between prices pl and p2 might be defined by l (Plp2) 1/(max(pl,p2)+I), to recognize that when it comes to customer interest, $5000 and $5020 are very similar, whereas $3 and $23 are not.) If the attribute is associative, then its value V may be decomposed as described above into a collection of real numbers, representing the association scores between the target object in question and various ancillary objects. V may therefore be regarded as a vector with components V1, V2, V3 etc., representing the association scores between the object and ancillary objects 1, 2, 3,	identical, i.e., “HAVE-Dsub-OCTOPUS”. A score for a document is illustratively a numeric sum of the weights of all uniquely matching triples in that document. All duplicate matching triples for any document are ignored. An illustrative ranking of the relative weightings of the different types of relations that can occur in a triple, in descending order from their largest to smallest weightings are: first, verb-object combinations (Dobj); verb-subject combinations (Dsub); prepositions and operators (e.g. Ops), and finally modifiers (e.g. Nadj).” Braden 25:41-48 “Rather than using fixed weights for each different attribute in a logical form triple, these weights can dynamically vary and, in fact, can be made adaptive. To accomplish this, a learning mechanism, such as, e.g., a Bayesian or neural network, could be appropriately incorporated into our inventive process to vary the numeric weight for each different logical form triple to an optimal value based upon learned experiences.” Culliss 10:47-52 “To present personalized search results to a particular person searching with a particular term or query, the present invention may display a number of articles from a number of the narrower related key term groupings or queries which are ranked by their respective previous-user relevancy scores.” Culliss 11:11-20 “It is also possible to consider both the previous-user relevancy

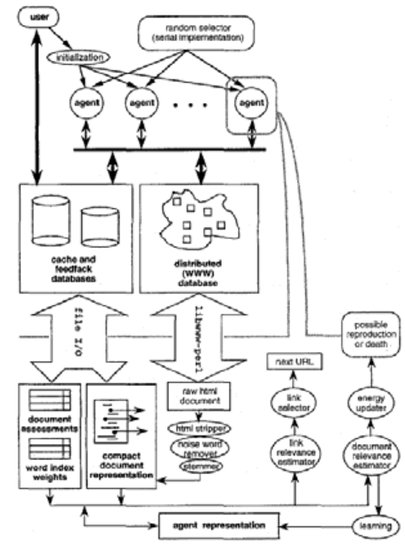
The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
		etc., respectively. The distance between two vector values V and U of an associative attribute is then computed using the angle distance measure, arccos (VU’/sqrt((Vv’)(UU’)). (Note that the three inner products in this expression have the form XY’=X1 Y1+X2 Y2+X3 Y3+... , and that for efficient computation, terms of the form Xi Y, may be omitted from this sum if either of the scores Xi and Y, is zero.) Finally, if the attribute is textual, then its value V may be decomposed as described above into a collection of real numbers, representing the scores of various word n-grams or character n-grams in the text. Then the value V may again be regarded as a vector, and the distance between two values is again defined via the angle distance measure. Other similarity metrics between two vectors, such as the dice measure, may be used instead.” Herz 1:25-28; 4:55-62 Herz contemplates using both “user profiles” and “query	score of the top narrower related key term groupings or queries, as well as the previous-user relevancy score of the articles under these narrower related key term groupings or queries. In this respect, the previous-user relevancy score of the top narrower related key term groupings or queries and the previous-user relevancy score of the articles under these narrower related key term groupings or queries can be combined in any possible manner, such as by adding, multiplying, or averaging together.” Culliss 5:18-21 “When a user first enters a search query, the personal data can be considered part of the request and stored within or added to the index, individually or in groupings with other items of data such as key terms, categories, or ratings.” Culliss 5:41-45 “When the next user enters a search request, the search request and the user’s personal data are combined to form groupings containing key term groupings, key terms and personal data groupings, category and personal data groupings, rating and personal data groupings, etc.” Culliss 10:8-13 “For example, when a woman enters the search request ‘shoes,’ the system can look for narrower related queries or key term groupings which contain or are related to the term ‘shoes’ and which have been entered by previous users having similar personal data, such as that of being a ‘woman.’” Ahn 3:43-46 “In step 414, the invention

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
		profiles” to form “target profile interest summaries” that “describe[] the user’s interest level in various types of target objects.” Herz 56:19-28 Herz further teaches that search profiles can be determined by “asking the user to specify search profiles directly by giving keywords and/or numeric attributes” (the search request/query profile) *and* by “using copies of the profiles of target objects or target clusters that the user indicates are representative of his or her interest” (the user profile). Herz 57:23-27 *Both* types of data are to be considered in determining which documents are most likely of interest to the user.	locates occurrences (hits) of the keyword in the document by traversing through the document’s document tree to find pertinent entries in the document’s document index table.” Dedrick *See, e.g.*, Figures 1-8, 8:20–9:24, 14:55–64. Krishnan 8:34-45 “The information access monitor IAM, at step 604, intercepts the query at step 603 and interprets the query. The information access monitor IAM, at step 604, uses the relevance index information stored in the index files IF to process the request and identify the ones of the objects previously indexed by document search engine DSE which match the relevance index information stored in index files IF.” Krishnan *See also* Fig. 6. Kupiec 4:60-63 “Verification is accomplished by lexico-syntactic analysis which looks for certain patterns in the user’s question and attempts to find corresponding or related patterns in documents.” Kupiec 10:41-46 “In one embodiment preliminary hypothesis generation comprises locating match sentences in the documents, scoring these match sentences, extracting noun phrases from the match sentences and from adjacent sentences in the primary documents, and scoring these noun phrases to generate a ranked list of preliminary hypotheses”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			Kupiec 14:45-53 “6.1 Lexico-Syntactic Analysis. Hypotheses are verified in step 260 through lexico-syntactic analysis. Lexico-syntactic analysis comprises analysis of linguistic relations implied by lexico-syntactic patterns in the input string, constructing or generating match templates based on these relations, instantiating the templates using particular hypotheses, and then attempting to match the instantiated templates, that is, to find primary or secondary documents that contain text in which a hypothesis occurs in the context of a template.” Kupiec 18:1-26 “6.5 Matching Templates Against Primary Documents. In step 264 an attempt is made to verify the linguistic relation under consideration for the hypothesis under consideration in the context of the primary documents. This is done by matching the filled-in templates generated in step 263 against the primary documents. In other words, sentences in which the hypothesis appears in the context of a template are sought in the primary documents. Any such sentences found are retained in association with the hypothesis as verification evidence for use in later processing steps. For example, if the template is “NP(Justice) (is, was) X” and the hypothesis is “Earl Warren,” the filled-in template is “NP(Justice) (is, was) Earl Warren,” and documents containing sentences such as “At that time the Chief Justice was Earl Warren . . . “ are potential matches. As another example, if the template is “X succeeded Shastri” and the hypothesis is “Indira Gandhi,” the filled-in template is

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			“Indira Gandhi succeeded Shastri.” The answer extraction subsystem seeks one or more primary documents that contain sentences conforming to this filled-in template, for example, “Indira Gandhi succeeded Shastri .... “ The testing of step 264 is carried out using only the primary documents. If sufficient template matches are found among the primary documents, then the linguistic relation is considered verified. In this case it is unnecessary to run secondary queries and steps 265 and 266 are skipped for this linguistic relation and hypothesis.” Reese 3:45-55 “The invention contemplates that the matching server 120 works with the client user profile request 100 to pare down the data delivered to the client. The matching server 120 pre-selects an aggregate of data that is determined to be the most relevant to different sets of user profile requests 100. The matching server 120 does this by searching various content sites 130, 140, 150, 160 on the Internet or other network. A user profile request 100 is applied against the matching server 120 aggregate of data like a sieve, and only data matching the user profile request 100 is returned to the client 110.” Menczer p. 159 “The user initially provides a list of keywords and a list of starting points, in the form of a bookmark file. In step (0), the population is initialized by pre-fetching the starting documents. Each agent is “positioned” at one of these document and given a random behavior (depending on the representation) and an initial reservoir of “energy”. In step

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			(2), each agent “senses” its local neighborhood by analyzing the text of the document where it is currently situated. This way, the relevance of all neighboring documents -those pointed to by the hyperlinks in the current document- is estimated. Based on these link relevance estimates, an agent “moves” by choosing and following one of the links from the current document.” Menczer p. 162 “Two agents born after 350 document have been visited and assessed, shown in Figures 7 and 8 respectively, have internalized some of the global environmental cues (d. Table 1) into their internal representations. Query words that are not very useful (e.g., INSTITUT and BRANCH) have disappeared from the keyword vectors through evolution, their places being taken by words that better correlate with user preferences (e.g., SYSTEM and PARTI). Menczer p. 160 “Figure 3: Architecture of the ARACHNID agent population.”

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			 Figure 3: Architecture of the ARACHNID agent population. Armstrong p. 4 “4. *Words used to define the user goal*. These features indicate words entered by the user while defining the information search goal. In our experiments, the only goals considered were searches for technical papers, for which the user could optionally enter the title, author, organization, etc. (see Figure 3). All words entered in this way throughout the training set were included (approximately 30 words, though the exact number varied with the training set used in the particular experiment). The encoding of the boolean feature in this case is assigned a 1 if and only if the word occurs in the user-specified goal and occurs in the hyperlink, sentence, or headings associated with this example.”

The '067 Patent	Salton '89	Herz	Additional Prior Art References
(g) calculating, by one of the local computer system and the remote computer system, a final match factor for each of said plural data item profiles, by adding said first similarity factor to at least one of said plural second similarity factors in accordance with at least one intersection between said first correlation and said second correlation;	Salton teaches calculating a final match factor. *See* p. 306, 313-9.	Herz 14:40-15:13 "Similarity Measures. What does it mean for two target objects to be similar? More precisely, how should one measure the degree of similarity? Many approaches are possible and any reasonable metric that can be computed over the set of target object profiles can be used, where target objects are considered to be similar if the distance between their profiles is small according to this metric. Thus, the following preferred embodiment of a target object similarity measurement system has many variations. First, define the distance between two values of a given attribute according to whether the attribute is a numeric, associative, or textual attribute. If the attribute is numeric, then the distance between two values of the attribute is the absolute value of the difference between the two values. (Other definitions are also possible: for example, the distance between prices pl and p2 might be defined by l (Plp2)	Salton '68 p. 414, Fig. 10-4. Braden 11:22-26 "Thereafter, through comparing the logical form triples for the query against those for each document, process 600 scores each document that contains at least one matching logical form triple, then ranks these particular documents based on their scores." Braden 17:44-53 "Of these triples, two are identical, i.e., "HAVE-Dsub-OCTOPUS". A score for a document is illustratively a numeric sum of the weights of all uniquely matching triples in that document. All duplicate matching triples for any document are ignored. An illustrative ranking of the relative weightings of the different types of relations that can occur in a triple, in descending order from their largest to smallest weightings are: first, verb-object combinations (Dobj); verb-subject combinations (Dsub); prepositions and operators (e.g. Ops), and finally modifiers (e.g. Nadj)." Braden 25:41-48 "Rather than using fixed weights for each different attribute in a logical form triple, these weights can dynamically vary and, in fact, can be made adaptive. To accomplish this, a learning mechanism, such as, e.g., a Bayesian or neural network, could be appropriately incorporated into our inventive process to vary the numeric weight for each different logical form triple to an optimal value based upon learned experiences."

The '067 Patent	Salton '89	Herz	Additional Prior Art References
		1/(max(pl,p2)+I), to recognize that when it comes to customer interest, $5000 and $5020 are very similar, whereas $3 and $23 are not.) If the attribute is associative, then its value V may be decomposed as described above into a collection of real numbers, representing the association scores between the target object in question and various ancillary objects. V may therefore be regarded as a vector with components V1, V2, V3 etc., representing the association scores between the object and ancillary objects 1, 2, 3, etc., respectively. The distance between two vector values V and U of an associative attribute is then computed using the angle distance measure, arccos (VU'/sqrt((Vv')(UU')). (Note that the three inner products in this expression have the form XY'=X1 Y1+X2 Y2+X3 Y3+... , and that for efficient computation, terms of the form Xi Y, may be omitted from this sum if either of the scores Xi and Y, is zero.) Finally, if the attribute is	Culliss 10:47-52 "To present personalized search results to a particular person searching with a particular term or query, the present invention may display a number of articles from a number of the narrower related key term groupings or queries which are ranked by their respective previous-user relevancy scores." Culliss 11:11-20 "It is also possible to consider both the previous-user relevancy score of the top narrower related key term groupings or queries, as well as the previous-user relevancy score of the articles under these narrower related key term groupings or queries. In this respect, the previous-user relevancy score of the top narrower related key term groupings or queries and the previous-user relevancy score of the articles under these narrower related key term groupings or queries can be combined in any possible manner, such as by adding, multiplying, or averaging together." Culliss 5:18-21 "When a user first enters a search query, the personal data can be considered part of the request and stored within or added to the index, individually or in groupings with other items of data such as key terms, categories, or ratings." Culliss 5:41-45 "When the next user enters a search request, the search request and the user's personal data are combined to form groupings containing key term groupings, key terms and personal data groupings, category and personal data groupings, rating and

The '067 Patent	Salton '89	Herz	Additional Prior Art References
		textual, then its value V may be decomposed as described above into a collection of real numbers, representing the scores of various word n-grams or character n-grams in the text. Then the value V may again be regarded as a vector, and the distance between two values is again defined via the angle distance measure. Other similarity metrics between two vectors, such as the dice measure, may be used instead." Herz 1:25-28; 4:55-62 Herz contemplates using both "user profiles" and "query profiles" to form "target profile interest summaries" that "describe[] the user's interest level in various types of target objects." Herz 56:19-28 Herz further teaches that search profiles can be determined by "asking the user to specify search profiles directly by giving keywords and/or numeric attributes" (the search request/query profile) *and* by "using copies of the profiles of target objects or target clusters that the user	personal data groupings, etc." Culliss 10:8-13 "For example, when a woman enters the search request 'shoes,' the system can look for narrower related queries or key term groupings which contain or are related to the term 'shoes' and which have been entered by previous users having similar personal data, such as that of being a 'woman.'" Culliss 7:44-63. Furthermore, Culliss contemplates determining the relevancy of a particular result to a particular query by considering *both* the relationship of the query to the user's personal data, *and* the relationship of a particular result to the user's personal data. Thus if a man inputs the query "shoes" he will get a different set of results than a woman who inputs the same query. Dedrick *See, e.g.*, Figures 1-8, 8:20–9:24, 14:55–64. Krishnan 8:34-45 "The information access monitor IAM, at step 604, intercepts the query at step 603 and interprets the query. The information access monitor IAM, at step 604, uses the relevance index information stored in the index files IF to process the request and identify the ones of the objects previously indexed by document search engine DSE which match the relevance index information stored in index files IF. This is accomplished by performing an object relevance determination based upon the identity of the user requesting the information, the user's profile and user's interest summary indexes stored in the database DB, and other user

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
		indicates are representative of his or her interest” (the user profile). Herz 57:23-27 *Both* types of data are to be considered in determining which documents are most likely of interest to the user.	profile criteria, administrative criteria, and object characterizing data.” Krishnan *See also* Fig. 6. Han p. 413 “One of the main tasks of the agent is to search the Web for documents that are related to the clusters of documents. The key question here is how to find a representative set of words that can be used in a Web search. With a single document, the words appearing in the document become a representative set. However, this set of words cannot be used directly in a search because it excessively restricts the set of documents to be searched. The logical choice for relaxing the search criteria is to select words that are very frequent in the document. The characteristic words of a cluster of documents are the ones that have high document frequency and high average text frequency. Document frequency of a word refers to the frequency of the word across documents. Text frequency of a word refers to word frequency within a document. We define the TF word list as the list of *k* words that have the highest average text frequency and the DF word list as the list of *k* words that have the highest document frequency. For each cluster, the word lists TF and DF are constructed. $TF \cap DF$ represents the characteristic set of words for the cluster, as it has the words that are frequent across the document and have high average frequency. The query can be formed as $(c_1 \wedge c_2 \ldots \wedge c_m) \wedge (t_1 \vee t_2 \ldots \vee t_n)$ where $c_1 = TF \cap DF$ and $t_1 = TF - DF$.”

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			Menczer p. 159 The user may assess any visited document D as relevant or non-relevant, with feedback $\phi(D) = \pm 1$. All the words in the document are also assessed by updating a "feedback list" of encountered words. Each word in this list, k, is associated with an integer count ω_k that is initialized with 0 and updated each time any document is assessed by the user: $\forall k \in D$ $$\omega_k \leftarrow \begin{cases} \omega_k + 1 & \text{if } \phi(D) = +1 \\ \omega_k - 1 & \text{if } \phi(D) = -1 \end{cases}$$ The word feedback list is maintained to keep a global profile of which words are relevant to the user. The output of the algorithm is a flux of links to document, ranked according to some relevance estimate—modulo relevance assessments by the user. Armstrong p.3 $$LinkUtility : Page \times Goal \times User \times Link \rightarrow [0, 1]$$ where $Page$ is the current web page, $Goal$ is the information sought by the user, $User$ is the identity of the user, and $Link$ is one of the hyperlinks found on $Page$. The value of $LinkUtility$ is the probability that following $Link$ from $Page$ leads along a shortest path to a page that satisfies the current $Goal$ for the current $User$. In the learning experiments reported here, we consider learning a simpler function for which training data is more readily available, and which is still of considerable practical use. This function is: $$UserChoice? : Page \times Goal \times Link \rightarrow [0, 1]$$ p.4 <table><tr><td>200 words Underlined</td><td>200 words Sentence</td><td>100 words Heading</td><td>≈ 30 words User goal</td></tr></table> Table 1: Encoding of selected information for a given $Page$, $Link$, and $Goal$. Where the value of $UserChoice?$ is the probability that an arbitrary user will select $Link$ given the current $Page$ and $Goal$. Notice here the $User$ is not an explicit input, and the function value predicts only whether users tend to select $Link$ – not whether it leads optimally toward to the goal. Notice also that information about the search trajectory by which the user arrived at the current page is not considered.

The '067 Patent	Salton '89	Herz	Additional Prior Art References
(h) selecting, by one of the local computer system and the remote computer system, one of said plural data items corresponding to a plural data item profile having a highest final match factor; and	Salton '89 p. 317-319 "Some of the advantages are the model's simplicity, the ease with which it accommodates weighted terms, and its provision of ranked retrieval output in decreasing order of query-document similarity."	Herz 57:24-27 "[T]he profile matching module 203 resident on proxy server S2 sequentially considers each search profile Pk from the user's search profile set to determine which news articles are most likely of interest to the user."	Salton '68 p. 12 "The results of a search performed with the Smart system appear as a ranked list of document citations in decreasing correlation order with the search request, as seen in the example of Fig. 1-6. The output of Fig. 1-6 is in a form suitable for communication with the user who originally submitted the search request." Braden 11:22-27 "Thereafter, through comparing the logical form triples for the query against those for each document, process 600 scores each document that contains at least one matching logical form triple, then ranks these particular documents based on their scores and finally instructs web browser 400 to present these particular documents, as symbolized by line 446." Culliss 3:19-25 "Demographic data includes, but is not limited to, items such as age, gender, geographic location, country, city, state, zip code, income level, height, weight, race, creed, religion, sexual orientation, political orientation, country of origin, education level, criminal history, or health. Psychographic data is any data about attitudes, values, lifestyles, and opinions derived from demographic or other data about users." Culliss 5:41-48 "When the next user enters a search request, the search request and the user's personal data are combined to form groupings containing key term groupings, key terms and personal data groupings, category and personal data groupings, rating and personal data groupings, etc. Articles

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			associated with these groupings are then retrieved from the index, and their relevancy scores are used or combined to determine their rankings." Dedrick *See, e.g.*, Figures 1-8, 22:49-53, 3:56 - 4:3, 8:20–9:24, 14:43–54, 16:23–32. Krishnan 5:1-9 "The information access monitor IAM then compares the object profiles with the users' interest summaries and user profiles to generate a rank ordered listing of objects most likely to be of interest to each user so that the information access monitor IAM can identify which information being retrieved via the gateway G is likely to be of interest to individual users from the plethora of objects available via the Internet I." Krishnan *See also* Fig. 6. Kupiec 5:16-18 "After all verification attempts are complete, the method rescores the hypotheses according to the degree to which they were successfully verified. In Example 1, Norman Mailer emerges as the winning answer hypothesis" Kupiec 10:59-64 "In step 280 the answer extraction subsystem performs hypothesis ranking according to a scoring scheme. The goal of this step is to rank highest the answer hypothesis or hypotheses most likely to be responsive to the input string. Step 280 is analyzed in more detail in section 5 below." Kupiec 21:22-32 "7.1 Scoring

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			In step 281 scores are assigned to the (unlinked) hypotheses. In one embodiment each hypothesis score is based on three criteria. The first criterion is verification evidence obtained through template matching in primary and secondary documents in step 260. The second criterion is co-occurrence of the hypothesis with phrases of the input string in primary and secondary documents, regardless of whether templates were matched. The third criterion is the preliminary hypothesis score developed in step 240, which is based on the scores of the primary document match sentences from which the hypothesis derives." Kupiec 25:18-20 "7.3 Ranking Hypotheses and Organizing Results In step 285 the hypotheses are ranked according to their scores from highest to lowest. This step can be accomplished by a straightforward sorting procedure." Menczer p. 159 The user may assess any visited document D as relevant or non-relevant, with feedback $\phi(D) = \pm 1$. All the words in the document are also assessed by updating a "feedback list" of encountered words. Each word in this list, k, is associated with an integer count ω_k that is initialized with 0 and updated each time any document is assessed by the user: $\forall k \in D$ $$\omega_k \leftarrow \begin{cases} \omega_k + 1 & \text{if } \phi(D) = +1 \\ \omega_k - 1 & \text{if } \phi(D) = -1 \end{cases}$$ The word feedback list is maintained to keep a global profile of which words are relevant to the user. The output of the algorithm is a flux of links to document, ranked according to some relevance estimate —modulo relevance assessments by the user.

The '067 Patent	Salton '89	Herz	Additional Prior Art References
(i) retrieving, by one of the local computer system and the remote computer system from the remote data storage system, said selected data item for display to the user, such that the user is presented with a data item having linguistic characteristics that substantially correspond to linguistic characteristics of the linguistic data generated by the user, whereby the linguistic characteristics of the data item correspond to the user's social, cultural, educational, economic background as well as to the user's psychological profile.	Salton '89 p. 229 "Information-retrieval systems process files of records and requests for information, and identify and retrieve from the files certain records in response to the information requests." Salton '89 p. 405-6 "To help furnish semantic interpretations outside specialized or restricted environments, the existence of a *knowledge base* is often postulated. Such a knowledge base classifies the principal entities or concepts of interest and specifies certain relationships between the entities. [43-45] . . . . The literature includes a wide variety of different knowledge representations . . . [one of the] best-known knowledge-representation techniques [is] the *semantic-net*. . . . In generating a semantic network, it is necessary to decide on a method of representation for each entity, and to relate or characterize the entities. The following types of knowledge representations are recognized: [46-48]. . . . A linguistic level in which the elements are language specific and the links represent arbitrary relationships between concepts that exist in the area under consideration." Salton '89 p. 409 "There is a substantial antinationalist tradition, however, which denies the idea of objective reality, and does not accept the existence off objects that bear properties independent of particular interpretations. [52-54] In this	Herz 58:27-34 "Once the profile correlation step is completed for a selected user or group of users, at step 1104 the profile processing module 203 stores a list of the identified articles for presentation to each user. At a user's request, the profile processing system 203 retrieves the generated list of relevant articles and presents this list of titles of the selected articles to the user, who can then select at step 1105 any article for viewing." Herz 66:65-67; 67:1-3 "The system uses the method of section 'Searching for Target Objects' above to automatically locate a small set of one or more clusters with profiles similar to the query profile, for example, the articles they contain are written at roughly an 8th-grade level and tend to mention Galileo and the Medicis."	Salton '68 p. 23 "Relations may exist between words that are not explicitly contained in the text but can be deduced from the context or from other texts previously analyzed; the identification of such relations requires deductive capabilities of considerable power." Braden 7:19-23 "Generally speaking and in accordance with our present invention, we have recognized that precision of a retrieval engine can be significantly enhanced by employing natural language processing to process, i.e., specifically filter and rank, the records, i.e., ultimately the documents, provided by a search engine used therein." Braden *See, e.g.*, 11:62-14:61. Culliss 3:19-25 "Demographic data includes, but is not limited to, items such as age, gender, geographic location, country, city, state, zip code, income level, height, weight, race, creed, religion, sexual orientation, political orientation, country of origin, education level, criminal history, or health. Psychographic data is any data about attitudes, values, lifestyles, and opinions derived from demographic or other data about users." Culliss 11:21-29 "When the previous-user relevancy score of the top narrower related key term groupings or queries is multiplied with the previous user-relevancy score of the articles under these narrower related key term groupings or queries for the search request of 'shoes' from a woman, for example, the following list of articles results . . . . These

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
	view, one cannot coherently talk about an external world without also furnishing the background and contexts that control the events in each circumstance.”		articles can then be presented to the woman user entering the search request ‘shoes’.” Dedrick 3:54–4:4 “The GUI may also have hidden fields relating to "consumer variables.” Consumer variables refer to demographic, psychographic and other profile information. Demographic information refers to the vital statistics of individuals, such as age, sex, income and marital status. Psychographic information refers to the lifestyle and behavioral characteristics of individuals, such as likes and dislikes, color preferences and personality traits that show consumer behavioral characteristics. Thus, the consumer variables refer to information such as marital status, color preferences, favorite sizes and shapes, preferred learning modes, employer, job title, mailing address, phone number, personal and business areas of interest, the willingness to participate in a survey, along with various lifestyle information. This information will be referred to as user profile data, and is stored on a consumer owned portable profile device such as a Flash memory-based PCMClA pluggable card.” Dedrick *See, e.g.*, Figures 1-8, 8:20–9:24, 14:43–54, 16:23–32. Krishnan 5:1-9 “The information access monitor IAM then compares the object profiles with the users’ interest summaries and user profiles to generate a rank ordered listing of objects most likely to be of interest to each user so that the information access monitor IAM can identify which information being

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			retrieved via the gateway G is likely to be of interest to individual users from the plethora of objects available via the Internet I." Krishnan *See also* Fig. 6. Kupiec 5:20-25 "Finally, the winning answer hypothesis can be presented to the user in conjunction with the documents and sentences in which it was found and the noun phrases that were used to verify it. In this way, the method shows not only what the answer is but why it was chosen." Kupiec 10:65-11:11 "In step 290 the answer extraction subsystem outputs a subset of the ordered list of answer hypotheses produced in step 280. The subset can be output directly to the user via the user interface. Alternatively or additionally it can stored in a storage device for later use, or made available for further processing. In some embodiments one or more answer hypotheses can be highlighted in the documents in which they appear for ease of reference. In other words, the answer extraction subsystem tells the user what it thinks the answer is and why. In some embodiments output to the user can be done in an interactive fashion, for example, by permitting the user to issue commands to the system to display answer hypotheses only, to display answer hypotheses in the context of the documents in which they appear, etc." Kupiec 25:53-26:10 "In step 287 the ranked hypotheses are organized into results suitable for output. In one embodiment in which results

The '067 Patent	Salton '89	Herz	Additional Prior Art References
			are to be presented to the user, the highest-ranked answer hypothesis is selected for presentation. This hypothesis is highlighted in the contexts in which it appears in primary and secondary documents, for example by displaying the document titles and the match sentences that confirm the linguistic relations implied by the user's question. The hypothesis can be emphasized through underlining or a distinctive font. Phrases of the input string that appear in context with the hypothesis can likewise be emphasized. Additionally, the answer extraction subsystem can provide further information about verification, linking, and scoring. In short, the answer extraction subsystem provides results that tell the user what the best answer hypothesis is, where it occurs in the documents, and why this answer was selected. The second and third-ranked hypotheses can be also presented, for example by themselves without the supporting information. In some embodiments, step 287 incorporates selecting which documents to present from numerous documents containing the best answer hypothesis. For example, if many documents match the best answer hypothesis, the one or two documents having the shortest matching sentences containing the hypothesis can be selected for presentation." Rapaport "For example, a particular user may be a nine-year-old child wanting to learn about butterflies" while another user maybe be "a post-graduate entomology student. Both users are interested in the same subject, but each desires different levels of sophistication in

The ‘067 Patent	Salton ’89	Herz	Additional Prior Art References
			information retrieval.” (1:32-38) Reese 4:51-53 “Other user profiles include, but are not limited to, areas of interest, business, politics, religion, education, etc.” Siefert teaches the use of “learning profiles,” which correspond to the user’s educational level, in order to return the correct resources to the user. (11:41-53). Han p.409: “WebACE submits the queries to the search mechanism and gathers the documents returned by the searches . . . . [T]he user can decide to add any or all of the new documents to his profile.” Menczer p. 159 “The output of the algorithm is a flux of links to document, ranked according to some relevance estimate – modulo relevance estimates by the user.”